A novel inhibitor protein for *Bombyx* cysteine proteinase is homologous to propeptide regions of cysteine proteinases

Yoshimi Yamamoto[a,*], Shoji Watabe[b], Takashi Kageyama[c], Susumu Y. Takahashi[a]

[a] *Department of Biochemistry and Radiation Biology, College of Agriculture, Yamaguchi University, Yamaguchi 753, Japan*
[b] *Radioisotopes Laboratory, Yamaguchi University, Yamaguchi 753, Japan*
[c] *Department of Biochemistry, Primate Research Institute, Kyoto University, Inuyama, Aichi 484, Japan*

Received 28 January 1999; received in revised form 10 March 1999

Abstract A cDNA clone for an inhibitor of *Bombyx* cysteine proteinase was isolated and sequenced. Active inhibitor proteins were expressed in *Escherichia coli* using the cDNA. The open reading frame of the cDNA encodes a 105 residues protein with 19 residues of a signal sequence. The inhibitor has amino acid sequences homologous to several cysteine proteinases, but only to their propeptide sequences. The results suggest that some cysteine proteinase proregions may have evolved as autonomous modules and become inhibitor proteins for cysteine proteinases.

Key words: Inhibitor; Cysteine proteinase; Propeptide region; cDNA cloning; Sequence homology

1. Introduction

Protein inhibitors of cysteine proteinases have been reported in various kinds of organisms and cystatins and the cystatin superfamily have been studied extensively [1–3]. A large amount of cysteine proteinase (BCP: *Bombyx* cysteine proteinase) which belongs to a papain superfamily accumulates as an enzymatically inactive pro-form in eggs and hemolymph of the silkmoth *Bombyx mori* [4,5]. Recently, we found and purified two forms of inhibitor proteins inhibiting the BCP (BCPI: BCP inhibitor), designated BCPI α and BCPI β, from the hemolymph of the silkmoth (Yamamoto et al., submitted). Both forms of inhibitor were inhibitory towards other cysteine proteinases such as papain, cathepsin L and B, but had no effects on trypsin and pepsin. Here, we isolated a cDNA clone for one of the inhibitor proteins and determined its nucleotide sequence. We also expressed the inhibitor in *Escherichia coli* and obtained active proteins. The deduced amino acid sequence of the inhibitor is significantly homologous to those of several cysteine proteinases, more precisely and most interestingly to the propeptide sequences of these enzymes. We present here a novel cysteine proteinase inhibitor which is homologous to the proregion of cysteine proteinases. There has been one report about proteins homologous to cysteine proteinase proregions [6]. Denizot, F. et al. have reported that two proteins, CTLA-2 α and CTLA-2 β, expressed in mouse activated T-cells and mast cells, are homologous to cysteine proteinase proregions. They have proposed that some cysteine proteinase proregions may have evolved as autonomous modules. The report presented here may be the second presentation which strongly supports their proposal.

2. Materials and methods

2.1. cDNA cloning

Routine molecular cloning techniques were used [7]. Inhibitor cDNAs were amplified using PCR techniques with the 'Marathon TM cDNA amplification kit (Clontech)' according to the user manual. mRNA was purified from fat bodies of the spinning *Bombyx* pupae. cDNA was synthesized and the specific adaptor DNAs supplied in the kit were ligated to both its ends. 5′ and 3′ rapid amplification of cDNA ends (RACE) reactions were performed with a combination of gene-specific primer and the adaptor primer using 'touchdown PCR' techniques suggested by the manual [8]. Two kinds of gene-specific primers (5′ RACE primer and 3′ RACE primer) were designed on the basis of the amino acid sequences obtained by amino acid sequencing of the purified inhibitor as shown in Fig. 1. To generate a full length cDNA, 5′ and 3′ cDNA clones were amplified by the second PCR with the adaptor primer and the corresponding gene-specific primer using the adaptor ligated cDNA. Gene-specific primers used for the second PCR (5′ cDNA primer and 3′ cDNA primer) were designed on the basis of the sequence of the 5′ and 3′ RACE clones which had been confirmed to be real by a nucleotide sequence analysis (Figs. 1 and 2). The nucleotide sequence analysis was performed by the dideoxy method of Sanger et al. [9], with a Hitachi Fluorescence DNA Sequencer (Hitachi, SQ-5500).

2.2. Expression of inhibitor in E. coli

cDNA constructs for expression in *E. coli* were amplified from an inhibitor cDNA covering the entire encoding region by PCR techniques. The primers used were designed on the basis of the cDNA sequence with additional restriction enzyme sites of *Bam*HI. The amplified DNA fragments were digested with *Bam*HI and ligated into pET20b (Novagen) cut with *Bam*HI. Ligated products were transformed first into *E. coli* XL1-Blue (Stratagene) and suitable clones were selected. Plasmids were then transferred to expression host strains *E. coli* BL21(DE3)pLysS. Expression was induced by isopropyl-β-D-thiogalactopyranoside (IPTG) and proteins from the periplasm were isolated with the osmotic shock procedure according to the 'pET System Manual (Novagen)'.

2.3. Other methods

The inhibition activity was measured with a procedure essentially the same as that for the BCP assay described previously [10]. The reaction mixture contained (in 300 μl) 100 mM sodium acetate buffer (pH 5.5), 1 mM EDTA, 8 mM cysteine, 10 μM Z-Phe-Arg-MCA (Peptide Institute), 30 ng of activated BCP and inhibitor proteins. After a short incubation (at 37°C for 5–10 min), the reaction was stopped and the remaining BCP activity was measured fluorometrically. Control mixtures containing no inhibitors (0% inhibition) were always taken in each assay. One unit of inhibitor activity was defined as the amount causing 50% inhibition of the BCP activity under the condition. SDS-PAGE and Western blotting analysis using a specific antibody against the inhibitor were carried out by as described previously in Section 2 [4,5]. A Northern blot analysis to determine the size of the inhibitor mRNA was carried out on poly (A) RNA, extracted from the fat bodies of the spinning pupae by hybridizing to a near full length inhibitor cDNA probe [5]. A sequence homology analysis of the GenBank data base was performed using the National

*Corresponding author. Fax: (81) (839) 33-5909.
E-mail: yamataka@po.yb.cc.yamaguchi-u.ac.jp

The nucleotide sequence presented here has been submitted to the EMBL database under the accession number of AJ131752

PII: S0014-5793(99)00382-8

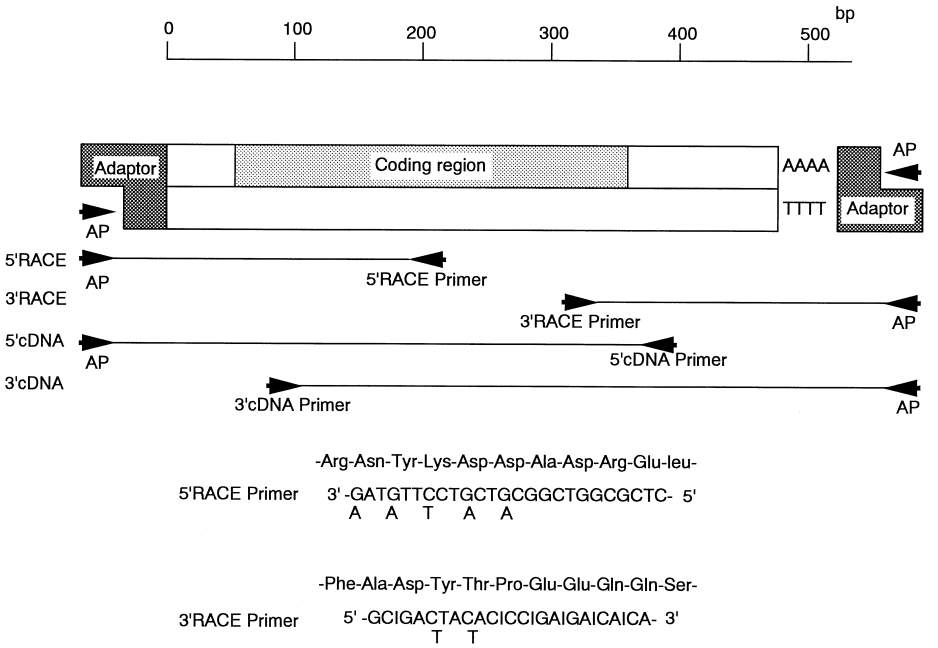

Fig. 1. Cloning strategy of cDNA clones of the BCP inhibitor. Synthesized cDNA was ligated with adaptor DNAs (black box) to create a population of cDNAs, from which specific cDNAs were amplified by PCR using different sets of gene-specific primers (5′ RACE primer, 3′ RACE primer, 5′ cDNA primer, 3′ cDNA primer) and adaptor primers (AP: 5′-CCATCCTAATACGACTCACTATAGGGC-3′). The light shadowed box indicates the amino acid encoding region.

Center for Biotechnology Information (NCBI) BLAST network service with the BLAST algorithm.

3. Results

3.1. Cloning and analysis of cDNA encoding inhibitor

After the first PCR amplification, 5′ and 3′ RACE products yielded two bands (1100 and 300 bp) and one band (350 bp) on agarose gel electrophoresis, respectively. At sequencing of the cloned cDNAs, a 5′ RACE clone (300 bp) and a 3′ RACE clone (350 bp) were identified with a significant homology to the amino acid sequences of the inhibitor. By the second PCR amplification following subcloning into the cloning vectors, two cDNA clones (5′ and 3′ cDNA) were isolated. The clones obtained and sequenced are presented in Fig. 1. To minimize sequencing errors caused by PCR artifacts (misreading by polymerase), at least three independent clones for each PCR product were isolated and sequenced in both orientations. Since the sequence of the 3′ end of the 5′ cDNA clone perfectly matched to the 5′ end sequence of the 3′ cDNA clone, it was possible to define a full length cDNA clone by combining the 5′ and 3′ cDNA clones (Fig. 2). The cDNA contained 522 nucleotides with a polyadenylation site (AATAAA) at 12 nucleotides upstream of the poly (A) tail. There was an open reading frame of 315 nucleotides encoding a 105 residue protein. Amino acid sequences of the N-terminal and Lys-C-digested peptides of the purified inhibitor matched well with the deduced amino acid sequence. The first 19 amino acid residues of the open reading frame seem to constitute the signal peptide according to the matching in the N-terminal amino acid sequence and theoretical considerations [11]. Two kinds of inhibitors (BCPI α and β) which are closely related but differ in the N-terminal amino acid sequences have been characterized in *B. mori* (Yamamoto et al., submitted). The deduced amino acid sequence indicates that the cDNA obtained here corresponds to the β form, the major form of the inhibitors. Northern blotting and hybridization using a cDNA fragment revealed the presence of a single mRNA of 580 bp for the inhibitor (Fig. 3).

3.2. Expression of a recombinant inhibitor in E.coli

In order to confirm that the protein encoded by the inhib-

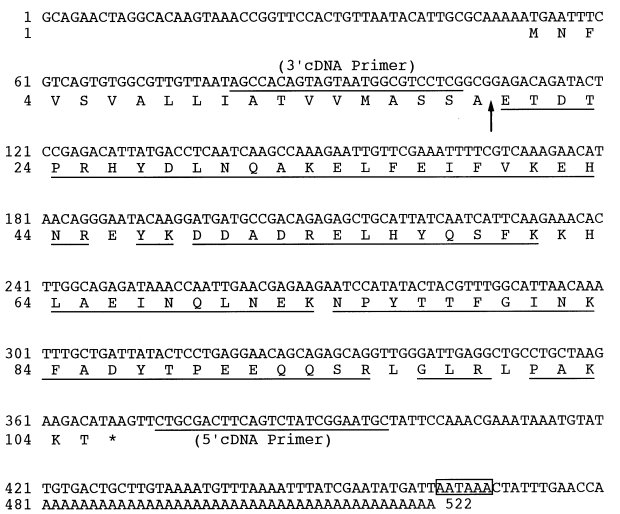

Fig. 2. The nucleotide sequence and deduced amino acid sequence of the BCP inhibitor. Amino acid sequences determined by protein sequencing are underlined. Nucleotide sequences shown by the underlinings were used to design primers for the second PCR. An arrow indicates a cleavage site in post-translational processing. A single asterisk indicates the termination codon and the polyadenylation signal is boxed.

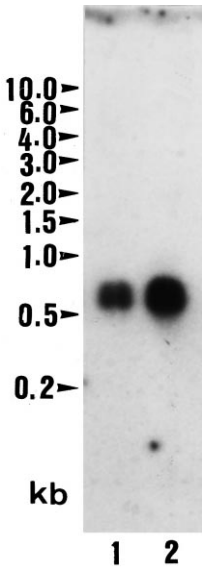

Fig. 3. Northern blot hybridization analysis. Poly (A) RNA (lane 1, 1 μg; lane 2, 2 μg) from fat bodies of spinning pupae were analyzed. A cDNA fragment (81–397) was used as a probe. Arrows indicate molecular weights estimated from a RNA marker (Novagen).

itor cDNA is really the BCP inhibitor, the protein was expressed in *E. coli*. We constructed a cDNA encoding the 86 amino acid residues (from Glu-20 to Thr-105) and lacking the amino acid sequence of the signal peptide of the inhibitor. The construct was ligated into an expression vector, in the site downstream of the sequence encoding the *E. coli* N-terminal signal sequence (*pelB* leader), which allowed the periplasmic localization of the fusion protein in *E. coli*. On an electrophoresis of the total protein from the *E. coli* transformed with a plasmid containing the inhibitor cDNA (pET20bBCPI), an additional protein band with an approximate molecular mass of 13 kDa was observed (Fig. 4a). The molecular

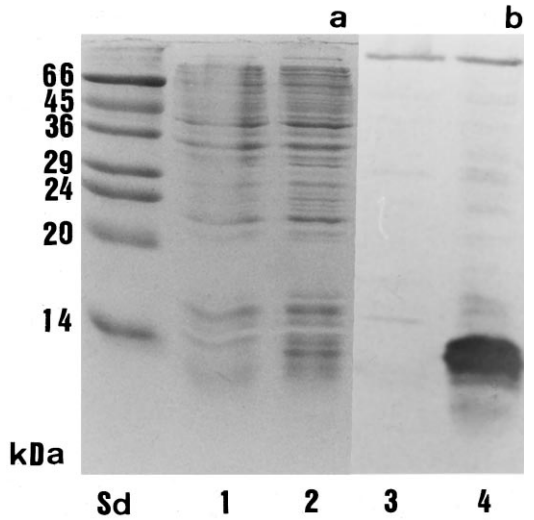

Fig. 4. SDS-PAGE analysis of BCP inhibitor expressed in *E. coli*. Total cell extracts from BL21(DE3)pLysSpET20b (lanes 1, 3) and BL21(DE3)pLysSpET20bBCPI (lanes 2, 4) were subjected to SDS-PAGE, following staining with Coomassie brilliant blue R250 (a) and immune staining with anti-BCP inhibitor serum (b). Sd, protein molecular weight standards.

mass coincides with the value estimated for the fused protein. A Western blotting analysis using specific antiserum against the inhibitor protein also indicates that the protein expressed is a fused protein with the inhibitor (Fig. 4b). Inhibitory activities against BCP were measured for extracts from the transformed cells. A significant activity (2870 U/mg protein) was detected in the extract containing the fused protein, whereas no activity was detected in the extract from cells with control vector (pET20b). Consequently, all of these data indicate that the cDNA we have presented and characterized corresponds to the inhibitor purified from *B. mori*.

3.3. Comparison of the inhibitor amino acid sequence with other proteins

A search in the GenBank sequence data base showed that the inhibitor contained sequence regions homologous to the propeptide sequences of cysteine proteinases (Fig. 5). The deduced amino acid sequence of the inhibitor was 28% identical (29/105 matches) to that of *Cydia pomonella* granulovirus cathepsin in the proregion [12] and showed a lower resemblance (23–25% identity) to the sequences of cathepsins from the other baculoviruses of *Choristoneura fumiferana* [13] and *B. mori* [14]. The amino acid sequence of the inhibitor

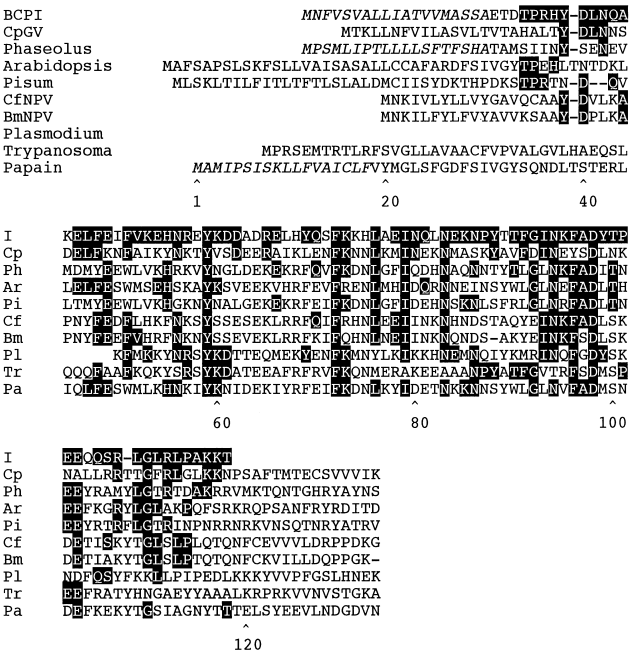

Fig. 5. Comparison of the amino acid sequence of BCP inhibitor with the propeptide sequences of related cysteine proteinases. The sequence of the BCP inhibitor was aligned with each of the other sequences by use of gaps to achieve maximal alignment. Numbers refer to the pre-propeptide sequence of papain. Residues that are identical to amino acid residues in the BCP inhibitor are indicated by black boxes and gaps are represented by dashes. Italics show signal peptides. BCPI, BCP inhibitor; CpGV, *Cydia pomonella* granulovirus cathepsin [12], g, AB010886; Phaseolus, *Phaseolus vulgaris* cysteine proteinase precursor, g, Z99952; Arabidopsis, *Arabidopsis thaliana* cysteine proteinase-like protein, g, AL022604; Pisum, *Pisum sativum* thiolprotease, g, X66061; CfNPV, *Choristoneura fumiferana* nucleopolyhedrovirus cathepsin L [13], g, P41715; BmNPV, *Bombyx mori* nucleopolyhedrovirus cysteine protease [14], g, U12688; Plasmodium, *Plasmodium fragile* cysteine proteinase [16], g, U33421; Trypanosoma, *Trypanosoma congolense* cysteine proteinase, g, S37048; Papain, *Carica papaya* papaya proteinase I [15], g, P00784; g, abbreviation of GenBank database accession number.

was also homologous to those of plant cysteine proteinases, with 27% identity for *Phaseolus vulgaris* cysteine proteinase, 28% identity for *Arabidopsis thaliana* cysteine proteinase-like protein, 27% identity for *Pisum sativum* thiolprotease and 21% identity for *Carica papaya* papain precursor [15]. The inhibitor was on average less homologous to protozoan cysteine proteinases, with 21% identity for *Plasmodium fragile* cysteine proteinase [16] and 18% identity for *Trypanosoma congolense* cysteine proteinase. Homology was not existent for the N-terminal ends which constitute signal peptides.

4. Discussion

Sequence comparison revealed that BCP inhibitor is not homologous to cystatin type inhibitors. The deduced amino acid sequence of the inhibitor was significantly homologous to the proregions of several cysteine proteinase precursors. The degree of identity is not high but almost the same identity has been observed when the proregions of cysteine proteinases are compared. The homology analysis of cysteine proteinases has revealed that the propeptide regions show a significant but low sequence homology, it amounts to 22–34% identity, although the mature regions of the proteinases show a very high sequence homology (30–60% identity) [17]. The protein sequence of CTLA-2 β is 36% identical to the proregion of mouse cathepsin L. The CTLA-2 β was expressed in the baculovirus system and the expressed protein has been biochemically characterized, showing that the protein is a competitive inhibitor of certain cysteine proteinases including papain, cathepsin L and H [18]. As shown in this report, the function of BCPI is undoubtedly inhibition of the cysteine proteinase activity. From these facts, we may be able to propose a new member of cysteine proteinase inhibitors, which resemble cysteine proteinases proregions but are distinct from cystatins.

The two CTLA-2 proteins (CTLA-2 α and β) are closely related at the amino acid sequences and the CTLA-2 genes are mapped to the same region of the mouse chromosome, suggesting that the ancestor gene of CTLA-2 has derived from an ancestor cysteine proteinase gene by partial duplication and duplicated again very recently to generate the two proteins. At present, we do not have data to suggest the origin of the inhibitor gene. But interestingly, for *B. mori*, there is at least one more inhibitor protein, BCPI α, which resembles the inhibitor cloned here (BCPI β) in the amino acid sequences. The N-terminal amino acid sequence of BCPI α is as follows: Asp-Thr-Asp-Pro-Pro^{5}-Arg-His-Tyr-Asp-Leu^{10}-Asn-Gln-Ala-Lys-Glu^{15}-Leu-Phe-Glu-Ser-Phe^{20}-Val-Lys. In the case of BCPI β, Asp-1, Pro-4 and Ser-19 are substituted by Glu-1, Thr-4 and Ile-19, respectively, other amino acid sequences determined to be identical for both inhibitors. All these results suggest that the inhibitor gene may have derived from an ancestor cysteine proteinase gene.

The function of the propeptides in the cysteine proteinases has been discussed [19,20]. Removal of the propeptide converts the inactive pro-forms of cysteine proteinases to the enzymatically active forms. Propeptides of procathepsin L [21–23] and papain precursor [24] inhibit selectively the enzyme activity of the corresponding mature enzymes. From these results, the possibility that one of the functions of the propeptide is inhibition of enzymatic activity has been proposed. The facts obtained here strongly support this possibility. The evolutional emergence of the inhibitor gene may be regarded as a genetic version of the functional propeptide. There are highly conserved amino acids interspersed in the sequences among the inhibitor and the propeptides of cysteine proteinases (Fig. 5). Similar highly conserved amino acids have been pointed out when the propeptide regions of cysteine proteinases are compared [17,25]. These results suggest that amino acid motives in which such highly conserved amino acids are interspersed serve to inhibit the proteinase activity. Further studies using recombinant inhibitor proteins are in progress.

References

[1] Barrett, A.J., Rawlings, N.D., Davies, M.E., Macheidt, W., Salvesen, G. and Turk, V. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G., Eds.), pp. 515–569, Elsevier Science Publishers BV, Amsterdam.
[2] Katunuma, N. and Kominami, E. (1995) in: Methods in Enzymology, Vol. 251, pp. 382–397, Academic Press, New York.
[3] Barrett, A.J. (1987) Trends Biochem. Sci. 12, 193–196.
[4] Kageyama, T. and Takahashi, S.Y. (1990) Eur. J. Biochem. 193, 203–210.
[5] Yamamoto, Y., Takimoto, K., Izumi, S., Toriyama-Sakurai, M., Kageyama, T. and Takahashi, S.Y. (1994) J. Biochem. 116, 1330–1335.
[6] Denizot, F., Brunet, J.-F., Roustan, P., Harper, K., Suzan, M., Luciani, M.-F., Mattel, M.-G. and Golstein, P. (1989) Eur. J. Immunol. 19, 631–635.
[7] Sambrook, J., Fritsch, F.F. and Maniatis, T. (1989) in: Molecular Cloning; A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, New York.
[8] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Proc. Natl. Acad. Sci. USA 85, 8998–9002.
[9] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[10] Takahashi, S.Y., Yamamoto, Y., Watabe, S. and Kageyama, T. (1997) Biochem. Mol. Biol. Int. 42, 591–600.
[11] von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[12] Kang, W., Tristem, M., Maeda, S., Crook, N.E. and O'Reilly, D.R. (1998) J. Gen. Virol. 79, 2283–2292.
[13] Hill, J.E., Kuzio, J. and Faulkner, P. (1995) Biochim. Biophys. Acta 1264, 275–278.
[14] Ohkawa, T., Majima, K. and Maeda, S. (1994) J. Virol. 68, 6619–6625.
[15] Cohen, L.W., Coghlan, V.M. and Dihel, L.C. (1986) Gene 48, 219–227.
[16] Rosenthal, P.J. (1996) Mol. Biochem. Parasitol. 75, 255–260.
[17] Ishidoh, K., Imajoh, S., Emori, Y., Ohno, S., Kawasaki, H., Minami, Y., Kominami, E., Katunuma, N. and Suzuki, K. (1987) FEBS Lett. 226, 33–37.
[18] Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E. and Muller, D. (1994) J. Biol. Chem. 269, 25172–25177.
[19] Tao, K., Stearns, N.A., Dong, J., Wu, Q. and Sahagian, G.G. (1994) Arch. Biochem. Biophys. 311, 19–27.
[20] Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C. and Mort, J.S. (1996) Structure 4, 405–416.
[21] Volkel, H., Kurz, U., Linder, J., Klumpp, S., Gnau, V., Jung, G. and Schultz, J.E. (1996) Eur. J. Biochem. 238, 198–206.
[22] Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J.S. and Menard, R. (1996) Biochemistry 35, 8149–8157.
[23] Jerala, R., Zerovnik, E., Kidric, J. and Turk, V. (1998) J. Biol. Chem. 273, 11498–11504.
[24] Taylor, M.A.J., Baker, K., Briggs, G.S., Connerton, I.F., Cummings, N.J., Pratt, K.A., Revell, D.F., Freedman, R.B. and Goodenough, P.W. (1995) Protein Eng. 8, 59–62.
[25] Karrer, K.M., Peiffer, S.L. and DiTomas, M.E. (1993) Proc. Natl. Acad. Sci. USA 90, 3063–3067.